单个率 Meta 分析中率的不同转换方法的比较

曾治宇

波科国际医疗贸易（上海）有限公司

【摘要】**目的** 在单个率 Meta 分析中对率的不同转换方法进行比较。**方法** 构造两套模拟数据进行单个率的 Meta 分析，考察 5 种数据转换方法（不转换、对数转换、logit 转换、平方根反正弦转换及双重反正弦转换）下的结果，兼顾固定效应模型和随机效应模型，及事件数为零时增加不同的固定值。计算汇总的率的均值（Mean），偏倚值（Bias）、偏倚率（Proportion Bias）、误差均方（Mean Squared Error, MSE）、误差均方百分比（Proportion MSE）及 95%可信区间的覆盖率（Coverage）。**结果** 对基于二项分布的单个率进行 Meta 分析时，平方根反正弦转换总体表现最佳。当事件数为零时，增加不同的固定值对结果影响较大，但这种校正对不转换的策略没有帮助，甚至有损；对于对数转换和 logit 转换的改善也非常有限。总体率<0.05 时，单个率 Meta 分析汇总的率偏倚较大。**结论** 单个率的 Meta 分析中平方根反正弦转换表现最佳。总体率<0.05 时使用 Meta 分析宜谨慎。

关键词 单个率；Meta 分析；数据转换；平方根反正弦转换

Comparison of Different Data Transformations for Meta Analysis of Single Proportions

ZENG Zhi-yu

Boston Scientific Corporation BSC Int'l Medical Trading (Shanghai) Co., Ltd. Beijing, 100020, China

[Abstract] Objective To compare different data transformations in Meta analysis for single proportions. **Methods** Two simulation data were constructed for Meta analysis under fixed effect model and random effect model, different adding values when event number was zero, and five different data transformation methods (raw proportion, log transformation, logit transformation, arcsine transformation and double arcsine transformation). Mean of pooled portion, bias, proportion bias, mean squared error, proportion mean squared error, proportion mean squared error and 95% confidence coverage were calculated. **Results** For Meta analysis of single proportions based on binomial distribution, generally, arcsine transformation performed best. When event number was zero, different values added to it did not improve the result much. Bias of pooled proportion was rather big when the population proportion is below 0.05. **Conclusion** Arcsine data transformation performed best during the simulation study for Meta analysis of single proportions. Caution should be used when dealing with population proportion less than

0.05.

Key words Single Proportions; Meta Analysis; Data Transformation; Arcsine Transformation

1 引言

Meta 分析有多种类型，单个率的 Meta 分析主要基于横断面研究的无对照二分类数据，这类数据常用于患病率、检出率、知晓率、病死率及感染率等的调查。这类数据的特点是仅有单组事件发生数和观察总数，而无对照组。本研究关注的是这类数据的 Meta 分析。

单个率 Meta 分析的数据对象属于计数资料。计数资料的描述包括比、比例和率等，但这些名称在临床研究的实际使用中往往容易混淆，一般来说，应注意比例（proportion，如患病率、治愈率等）和率（rate 或 incidence，发生率，如死亡率、发病率等）的区别。本研究不包括以时间事件数据表达的率，也不包括诊断试验中的敏感性、特异性等效应量。

单个率 Meta 分析通常采用倒方差法进行合并，因此在样本量或率较小时，常需要进行适当的数据转换。许多软件提供了不同的数据转换方法，如不转换（直接使用原始率）、对数转换、logit 转换、平方根反正弦转换、双重反正弦转换等。但在实际操作中具体应在什么情况下转换，首选何种转换方法并不明确。本研究拟通过模拟分析，比较单个率 Meta 分析不同方法数据转换的结果，为这一类型的 Meta 分析的方法选用提供理论依据。

2 方法

2.1 模拟数据

模拟构建 2 套用于单个率 Meta 分析的数据，验证 5 种率的转换方法的结果。

模拟数据 1 为参考既往文献的模拟数据[1]，假定有 9 个研究，样本量从 20 开始，以 20 递增直至 180。假设每个研究中的事件数符合既定的样本量与总体率（如患病率）的二项分布。在固定效应模型中，假设患病率真值是固定的 P_0，分别为 0.01、0.05 和 0.2。对于随机效应模型，考虑 P 服从均数为 P_0，标准差分别为 $0.1P_0$、$0.2P_0$、$0.5P_0$ 及 P_0 的正态分布。这样总共产生 3+3*3=12 种情形。对每种情形随机抽样 1000 次，对每一次抽样进行一组 Meta 分析，包括 5 种转换方法，对于事件数为零时，考察数值增加 0、 0.1、0.2、0.3……0.9 及 1 等情况下的结果，这样总计 12*5*11*1000=660,000 次 Meta 分析。

模拟数据 2 为医疗并发症模拟数据。假定有 5 个研究，样本量分别为 50、100、200、300、1000（1 个小样本，3 个中等样本，1 个大样本），只考虑固定效应模型，并发症率从 0.01 至 0.2，按 0.01 递增，共 20 种情况。每种条件下抽样 1000 次，每一次抽样进行一组 Meta 分析，包括 5 种转换方法，事件数为零时增加 0.5 或不变，这样总计 20*5*2*1000=200,000 次 Meta 分析。

2.2 单个率 Meta 分析方法

统计软件使用 R（版本 3.5.2），利用 R 语言 Meta 包的 Metaprop 函数，根据 5 种数据转换方法及不同的零事件处理方式进行单个率的 Meta 分析。对每种条件下的 1000 次 Meta 分析，计算汇总的率的均值（Mean），偏倚值（Bias）、偏倚率（Proportion Bias）、误差均方（Mean Squared Error, MSE）、误差均方百分比（Proportion MSE）及 95%可信区间的覆盖率（Coverage），具体方法见以下公式。

因正态近似法构建的 95%可信区间，在率较小的时候对不转换的情况显然不利，为避免研究偏倚，95%可信区间的计算一律采用 Clopper-Pearson 方法。

3 结果

3.1 模拟数据 1 结果

图 1 给出了固定效应模型的结果（P_0=0.05），横坐标是事件数为零时的增加值（由 0 至 1，按 0.1 递增）。由 A 栏可以看出，汇总率的均值，以绿色的平方根反正弦转换和蓝色的双重反正弦转换最接近 0.05。原始率的结果较差，事件数为零时的增加值对汇总率有较明显的影响，但校正的结果并不满意。相应的，B 栏是偏倚百分比，仍然是平方根反正弦法偏倚率的绝对值最小，双重反正弦转换、logit 和对数转换紧随其后。C 栏是误差均方百分比，D 栏是 95%可信区间的覆盖率，在这两个参数上 4 种转换方法的结果都非常接近，明显优于不转换的情况。表 1 列出了事件数为零时数值不变或增加 0.5 的两种情况的具体数值。

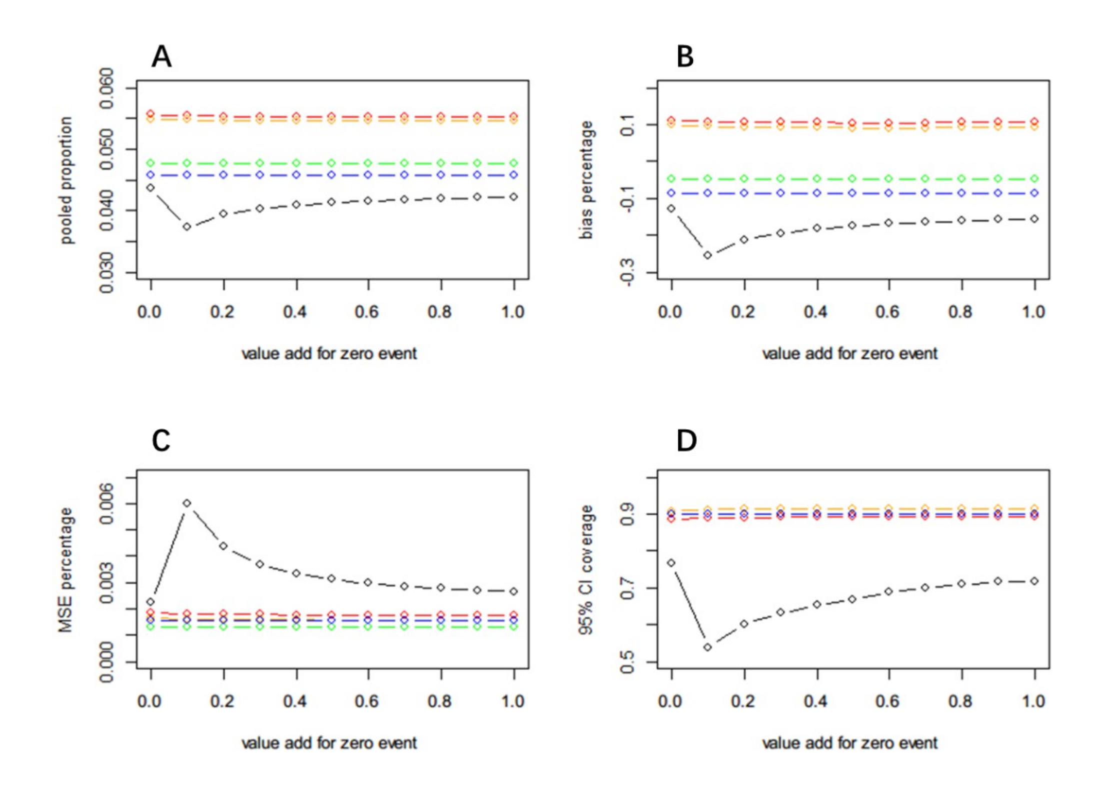


图 1 固定效应模型结果 P_0=0.05

横坐标均为当事件数为零时的增加值。A 为汇总率的均值，B 为偏倚率，C 为误差均方百分比，D 为 95%可信区间覆盖率。黑色为原始率，红色为对数转换，橙色为 logit 转换，绿色为平方根反正弦转换，蓝色为双重反正弦转换。

表 1 单个率 Meta 分析固定效应模型 P_0=0.05

增加值	转换	均值	偏倚率	误差均方百分比	覆盖率
0	原始率	0.043591	-0.128183	0.0022312	0.766
	对数	0.055594	0.111872	0.0018431	0.885
	logit	0.054915	0.098300	0.0016486	0.908
	平方根反正弦	0.047652	-0.046956	0.0013097	0.900
	双重反正弦	0.045653	-0.086935	0.0015500	0.898
0.5	原始率	0.041248	-0.175034	0.0031139	0.668
	对数	0.055243	0.104857	0.0017666	0.893
	logit	0.054543	0.090865	0.0015768	0.914
	平方根反正弦	0.047652	-0.046956	0.0013097	0.900
	双重反正弦	0.045653	-0.086935	0.0015500	0.898

类似的，图 2 是随机效应模型的结果（P=0.05，SD=0.005），由 A 栏可以看出汇总率的均值，以绿色的平方根反正弦转换和蓝色的双重反正弦转换最接近 0.05，原始率的结果较差，事件数为零时的增加值对汇总率有较明显的影响，但校正的结果并不满意。对数转换和 logit 转换的结果也较差。相应的，B 栏是偏倚百分比，仍然是平方根反正弦和双重反正弦转换的偏倚率的绝对值最小，原始率、logit 和对数转换的结果较差。C 栏是误差均方百分比，平方根反正弦和双重反正弦转换最优。D 栏是 95%可信区间的覆盖率，平方根反

正弦转换和双重反正弦最优且几乎重叠。表 2 列出了事件数为零时数值不变或增加 0.5 的两种情况的具体数值。另外，在随机效应模型中，随着并发症率 P 的标准差的增加，汇总率的偏倚率和覆盖率也逐渐变差（结果未呈现）。

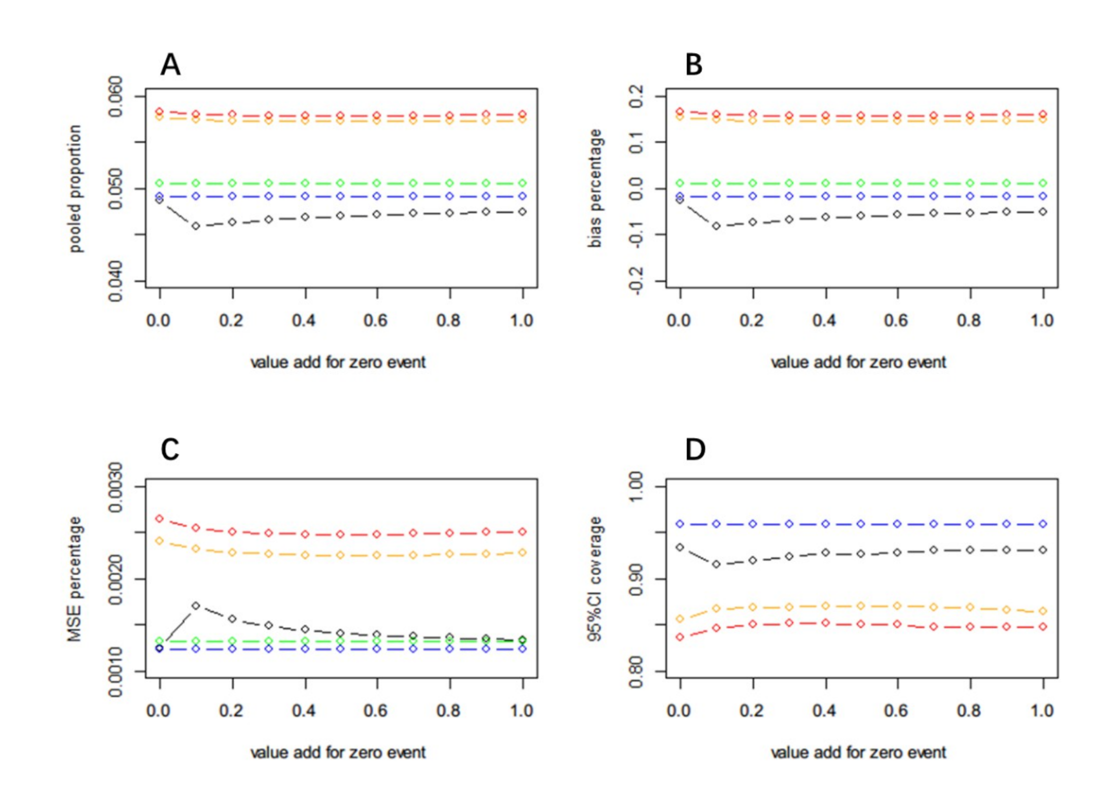


图 2 随机效应模型 P=0.05, SD=0.005

A 为汇总率的均值，B 为偏倚率，C 为误差均方百分比，D 为 95%可信区间覆盖率。黑色为原始率，红色为对数转换，橙色为 logit 转换，绿色为平方根反正弦转换，蓝色为双重反正弦转换。

表 2 单个率 Meta 分析随机效应模型 P=0.05,SD=0.005

增加值	转换	均值	偏倚率	误差均方百分比	覆盖率
0	原始率	0.048657	-0.026854	0.0012430	0.933
	对数	0.058286	0.165720	0.0026361	0.836
	logit	0.057686	0.153725	0.0024035	0.855
	平方根反正弦	0.050537	0.010749	0.0013151	0.958
	双重反正弦	0.049107	-0.017867	0.0012338	0.959
0.5	原始率	0.046953	-0.060942	0.0014082	0.926
	对数	0.057821	0.156419	0.0024752	0.850
	logit	0.057206	0.144113	0.0022496	0.870
	平方根反正弦	0.050537	0.010749	0.0013151	0.958
	双重反正弦	0.049107	-0.017867	0.0012338	0.959

3.2 模拟数据 2 结果

图 3 显示了模拟数据 2 即假想并发症数据的结果，横坐标为并发症的总体率（从 0.01 至 0.2，按 0.01 递增）。A 栏和 B 栏都是偏倚率，A 栏为当事件数为零时保持不变，B 栏则增加 0.5。由 A 栏可以看出在总体率<0.05 时，各方法的偏倚率都较大，但偏倚率随着总体

率的增加而迅速减小。对数转换和 logit 转换高估了并发症率，而不转换、平方根反正弦和双重反正弦转换均低估了并发症率。总体上绿色的平方根反正弦转换的偏倚率的绝对值最小。B 栏显示增加 0.5 后偏倚率改善不明显，且原始率的偏倚率反而变差。C 栏为误差均方百分比，D 栏为 95%可信区间的覆盖率，平方根反正弦、对数及 logit 转换后的覆盖率较高。

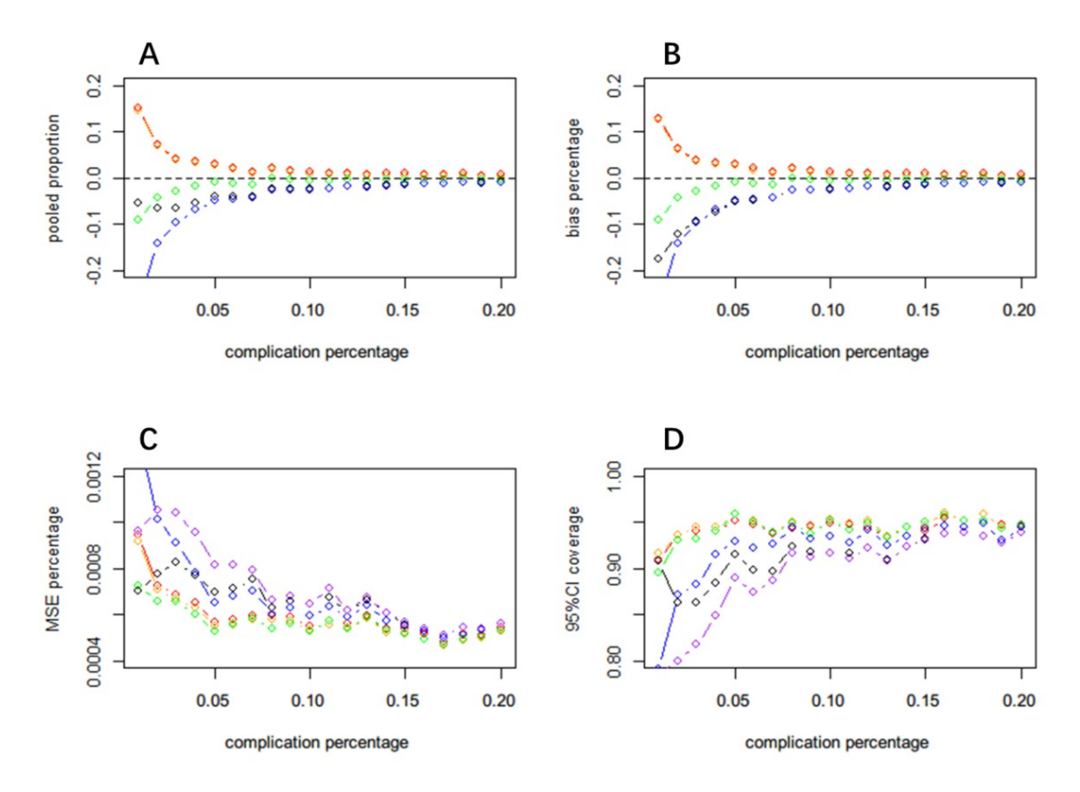


图 3 并发症模拟数据 Meta 分析的结果

横坐标为并发症的总体率，从 0.01 至 0.2，按 0.01 递增。A、B 均为汇总率的均值，当事件数为零时，A 不变，B 增加 0.5。C 为误差均方百分比，D 为 95%可信区间覆盖率，黑色为原始率，红色为对数转换，橙色为 logit 转换，绿色为平方根反正弦转换，蓝色为双重反正弦转换，紫色为当事件数为零时，原始率的事件数增加 0.5。

4 讨论

本研究通过模拟分析，发现对基于二项分布的单个率进行 Meta 分析，平方根反正弦转换总体表现最佳。当事件数为零时，增加不同的固定值对结果影响较大，但这种校正对不转换的策略没有帮助，甚至有损；对对数转换和 logit 转换改善非常有限。事件数为零时不影响平方根反正弦和双重反正弦的结果。总体率小于 0.05 时，Meta 分析汇总的率偏倚较大。

单个率的 Meta 分析采用倒方差法合并时，率较小时获得的权重过大。此时应考虑率的不同的转换方法。根据数据的特点和转换的目的，数据转换可有不同的方法[2]，而用于单个率 Meta 分析的率的转换的常见方法有以下 5 种：1）不转换（原始率直接计算），2）对数转换，3）logit 转换，4）平方根反正弦转换及 5）双重反正弦转换（double arcsine transformation，又称 Freeman-Tukey transformation）[1, 3-5]，这些方法多已纳入 Meta 分析的统计软件中。

但具体操作时何时选用哪种转换方法尚无定论。国内有一些单个率 Meta 分析软件使用的介绍[6-9]，检索国外文献，类似研究也不多见[1, 4]。Trikalinos 等[4]对单个比例和发生率的 Meta 分析分别进行了进行模拟研究，在单个比例的 Meta 分析中，根据比例的分布、均值、方差、研究个数及研究样本量等构建了 700 余种 Meta 分析的情形，对每种情形进行 1000 次随机抽样模拟，同时考察固定效应模型和随机效应模型。 率的转化包括了不转换、logit 转换和平方根反正弦转换等 3 种方法，将 Meta 分析得到的率与假设的真值率进行比较，比较的指标包括差值、差值百分比、均方差、均方差的百分比、以及 95%可信区间（单个分析以 Wald 法计算，再汇总 1000 个分析）覆盖真值的比例，结果显示平方根反正弦转换是较好的方法。本研究的结果与之一致。Barendregt 等[1]对患病率（Prevalence）的 Meta 分析进行了模拟研究，假设了样本量为 20-180 的 9 个研究，设定患病率为 0.05，随机效应模型时假设患病率在不同的研究中为正态分布，均值为 0.05，标准差为 0.005。进行 1000 次随机抽样，率的转换方法包括了不转换、logit 转换及双重反正弦转换等 3 种方法。该研究比较了 Meta 分析得到的值与假设的真值的偏差、均方差及 95%可信区间（也是 Wald 法计算）覆盖真值的比例，结果双重反正弦法略优于 logit 转换（如固定效应模型时的偏倚率在原始率、logit 转换及双重反正弦转换中分别为-0.022，0.004，0.002）。本研究在模拟数据 1 部分的结果与之相似，在固定效应模型中，两者接近；而在随机效应模型中双重反正弦转换优于 logit 转换。在模拟数据 2 部分的结果中，双重反正弦转换与 logit 转换的结果也较为接近，且 logit 转换在误差均方百分比及覆盖率指标上明显更优，但都不如平方根反正弦转换。

上述的两个研究各自都只比较了 3 种率的转换方法，且没有平方根反正弦转换与双重反正弦转换的直接对比。而且 Barendregt 等[1]的研究假设情形较为单一。本研究比较了 5 种率的转换方法，并且在两套模拟数据中都显示平方根反正弦转换最优。

本研究的主要局限在于：1）统计软件只选择了 R，不一定能代表其他的软件。R 软件作为一款免费的软件，功能强大而灵活，应用越来越广泛，且当前使用 R 作 Meta 分析的研究也越来越多。2）模拟分析的情景仍不够丰富。但本研究设计了两套模拟数据，考察了事件数为零时增加不同的固定值，及总体率从 0.01 至 0.2 的情况，应当具有一定的代表性。

总之，本研究认为单个率 Meta 分析，宜选用平方根反正弦转换，而事件数为零时增加一固定值的做法并不能明显改善结果。

5 参考文献

[1] Barendregt J J, Doi S A, Lee Y Y, et al. Meta-analysis of prevalence[J]. Journal of Epidemiology and Community Health,2013,67(11):974-978.
[2] 孙振球，徐勇勇. 医学统计学[M]. 第 4 版. 北京:人民卫生出版社,2014.
[3] Nyaga V N, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data[J]. Arch Public Health,2014,72(1):39.

[4] Trikalinos T A, Trow P, Schmid C H. Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates[M]. Rockville (MD):Agency for Healthcare Research and Quality (US),2013.
[5] Schwarzer G. General Package for Meta-Analysis[Z]. 2017:2018.
[6] 王佩鑫，李宏田，刘建蒙. 无对照二分类资料的 Meta 分析方法及 Stata 实现[J]. 循证医学,2012,12(01):52-55.
[7] 罗美玲，谭红专，周权，等. 在 R 软件中实现单个率的 Meta 分析[J]. 循证医学,2013,13(03):181-184.
[8] 肖静，吴桂云，高月霞，等. 单一率 Meta 分析的 Matlab 软件实现[J]. 中国卫生统计,2014,31(04):715-717.
[9] 陈月红，杜亮，耿兴远，等. 无对照二分类数据的 Meta 分析在 RevMan 软件中的实现[J]. 中国循证医学杂志,2014,14(07):889-896.